# COMPOSABLE SIMULATION RESEARCH AT THE AIR FORCE INSTITUTE OF TECHNOLOGY

**MAJOR KARL S. MATHIAS**

**Deputy Head, Department of Electrical and Computer Engineering**

# Introduction

- Problem: Need technology for rapidly specifying and building simulation scenarios
- Approach: Investigate visual languages for specification of simulation scenarios, develop advanced, intelligent component repositories
- Results:
  - Prototype extensions to UML that allow specification of simulation scenarios
  - Created repository that uses modern information retrieval techniques to store/extract scenario components

# Overview

- The Air Force Institute of Technology
- Visual Languages for Simulation Modeling
  - Problems with Representation
  - What Makes a Good VL for Composable Sim?
  - Results/Future Work
- Intelligent Scenario Component Repositories
  - Problems with Building Scenarios
  - Using Info Retrieval Engines/Parser Generators/XML
  - Results/Future Work

# AFIT

- The Air Force’s Graduate School
- Graduate 150 – 200 Students Per Year
  - Master’s, PhD Programs
  - EE, Comp Sci, Ops Research, Physics, Aero, Environmental Research, Info Sys Mgt, etc.
  - Officers: Average rank O-2
  - Enlisted: Starting this fall, average rank E-7
- Focused on DoD, AF Research Problems
- Partnered with Industry when Possible

# Visual Languages

# Visual Languages

- Object-Oriented Analysis Design Languages
  - Can be traced back to the 1970's (Chen's ER diagrams)
  - By early 1990's, over 50 methods available
  - Major players: Booch, Rumbaugh, Jacobs, Coad & Yourdan
- Unified Modeling Language
  - Unification of Booch, Rumbaugh, Jacobson under one methodology
  - Structure, Behavior, Use Cases under one roof
  - Aimed at the world, but used almost exclusively for software engineering

# What Makes a VL Good?

- Expressiveness – Can it express everything needed?
- Frequency of Errors – Does it avoid user mistakes?
- Redundancy – Does it avoid redundant specifications?
- Locality of Change – Do changes propagate and cause undesirable side-effects?
- Reusability – Does it promote component reuse?
- Reliability – Does it promote model consistency?
- Translatability – Can it be translated into a simulation?
- Compatibility – Does it favor one specific sim domain?

# Why Not UML?

- Appears to be an 80% solution
  - Excellent ability to specify static architectures
  - Some behavior specification (interaction,S-T diagrams)
- Scenario specifications are UML “instance diagrams”
  - UML has limited facilities to support instance diagrams
  - Behavior must be thought of abstractly
  - Temporal behavior difficult to specify (“Unit 23 arrives at 1600 hrs”)
  - Scalability issues – can’t design a component and then use it as a “black box” in other designs
  - Oriented toward engineers – not analysts

# Approach

- Common Design Pattern: Most sim domains are composed of:
  - Components: (Logic Gates, Queues, Aircraft, Weapon)
  - Relations: (Cables, Links, Command Chain)
- Analysts:
  - Browse Components
  - Instantiate into Environment
  - Connect to Other Components
  - Invoke Model Consistency Check
  - Translate into Simulation Scenario Syntax

# VL Elements

A_DataType

AnInt:int
AString:String

- Analogous to “class”
- Grouping and Abstraction Mechanism

A_PortType

ADouble:double
ALong:long

- A special type of “class” that specifies a port type
- Interface of component

A_RelationType

AChar:char
ABoolean:boolean

- Specifies a relation type in the modeling domain

A_ComponentType

AFloat:float

APortInstance:A_PortType
AnotherOne:A_PortType

- Specifies a component type in the modeling domain

# VL Elements

Type Extension

- Analogous to “inheritance”

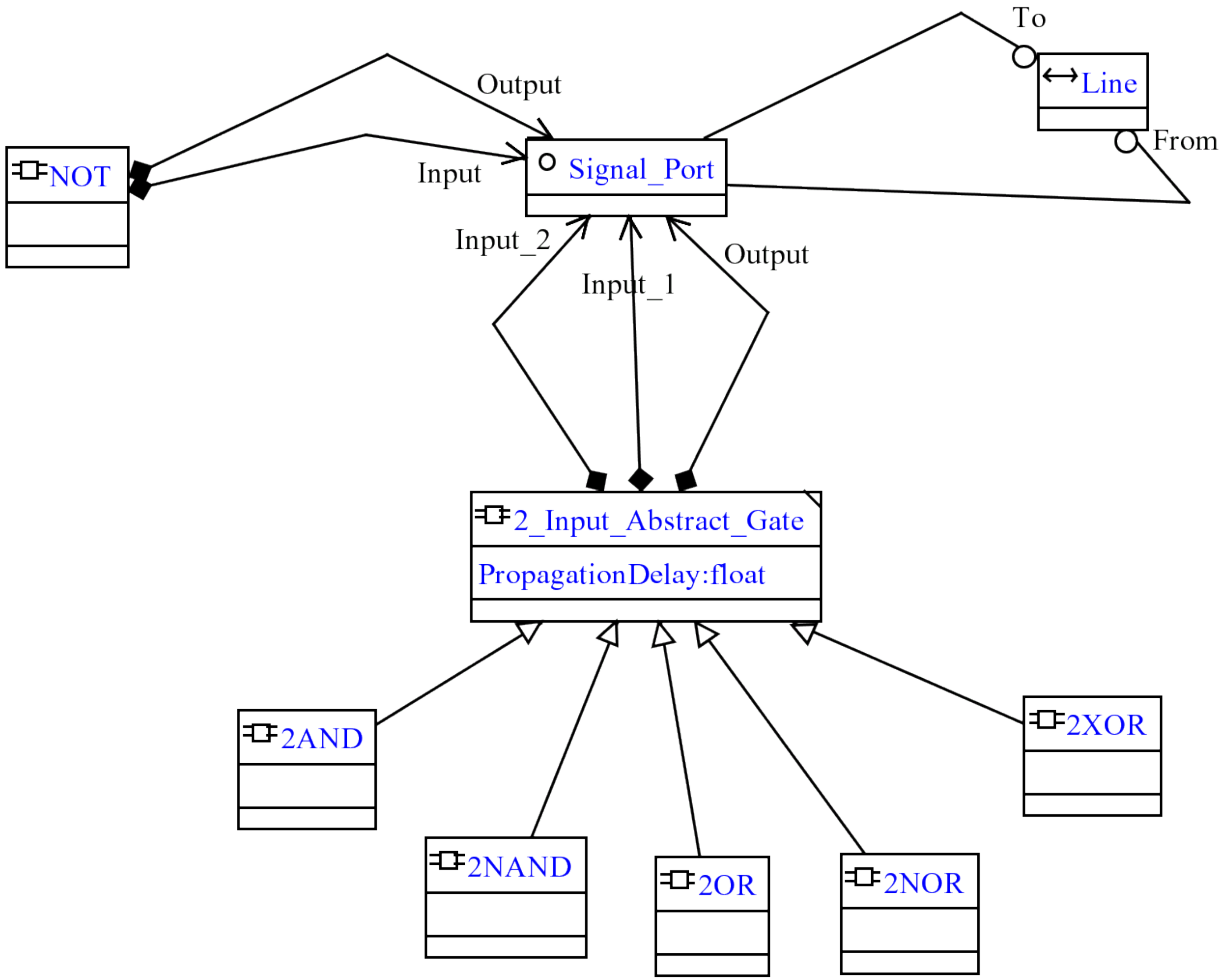
Composition

- Analogous to “composition”
- Strong Coupling

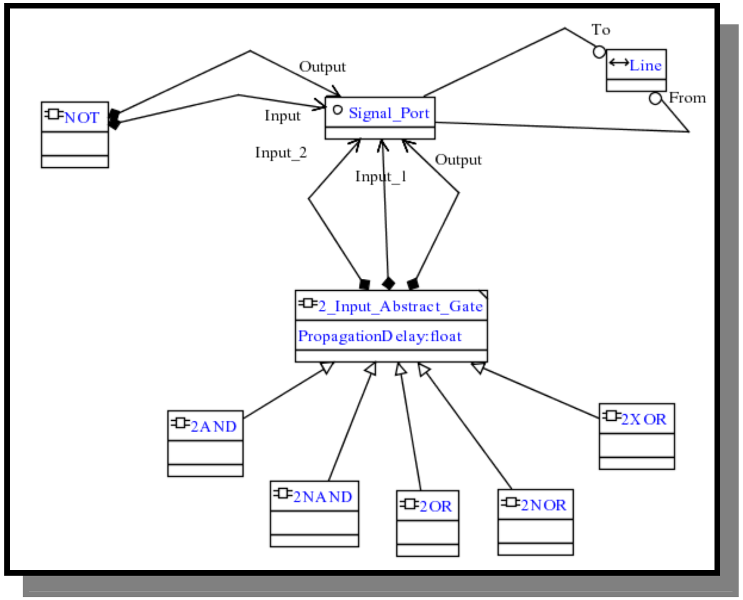
Port Type Selection

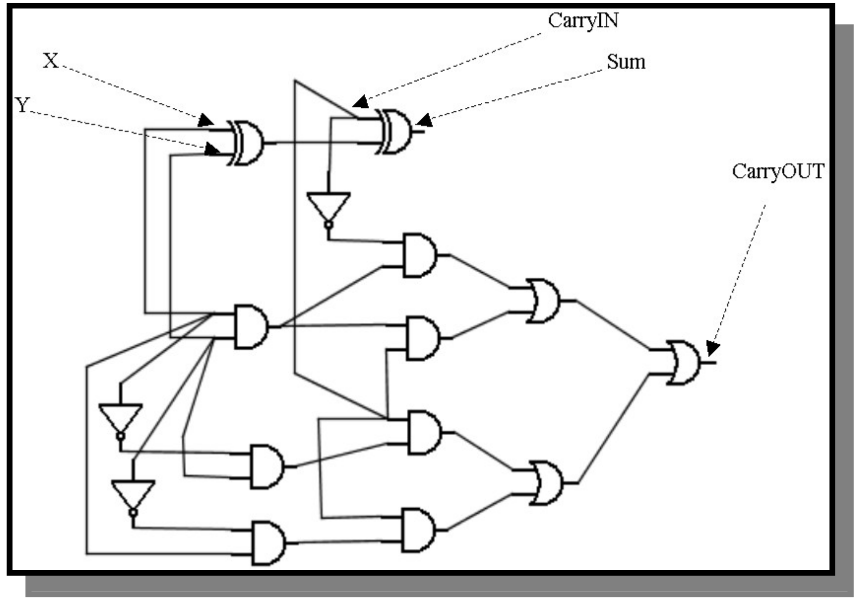
- Visualizes the port types that a relation type connects

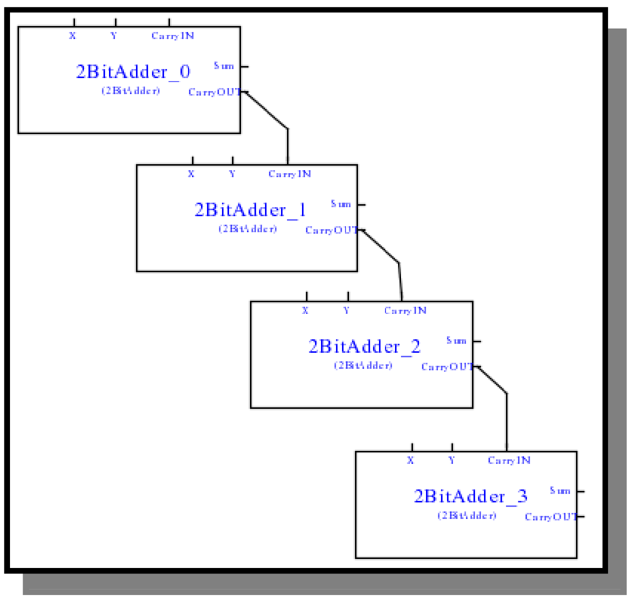
# Low-Level Design

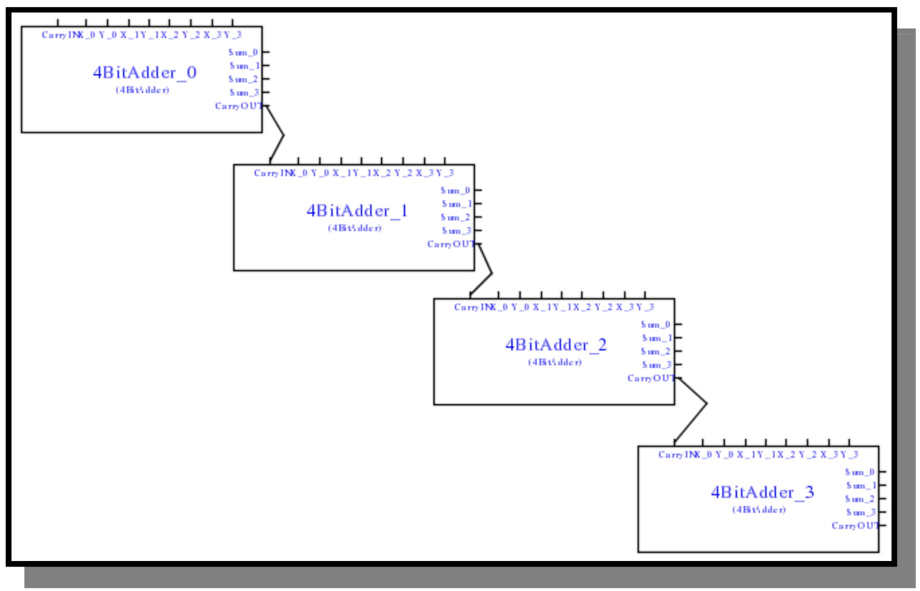


# Component Design

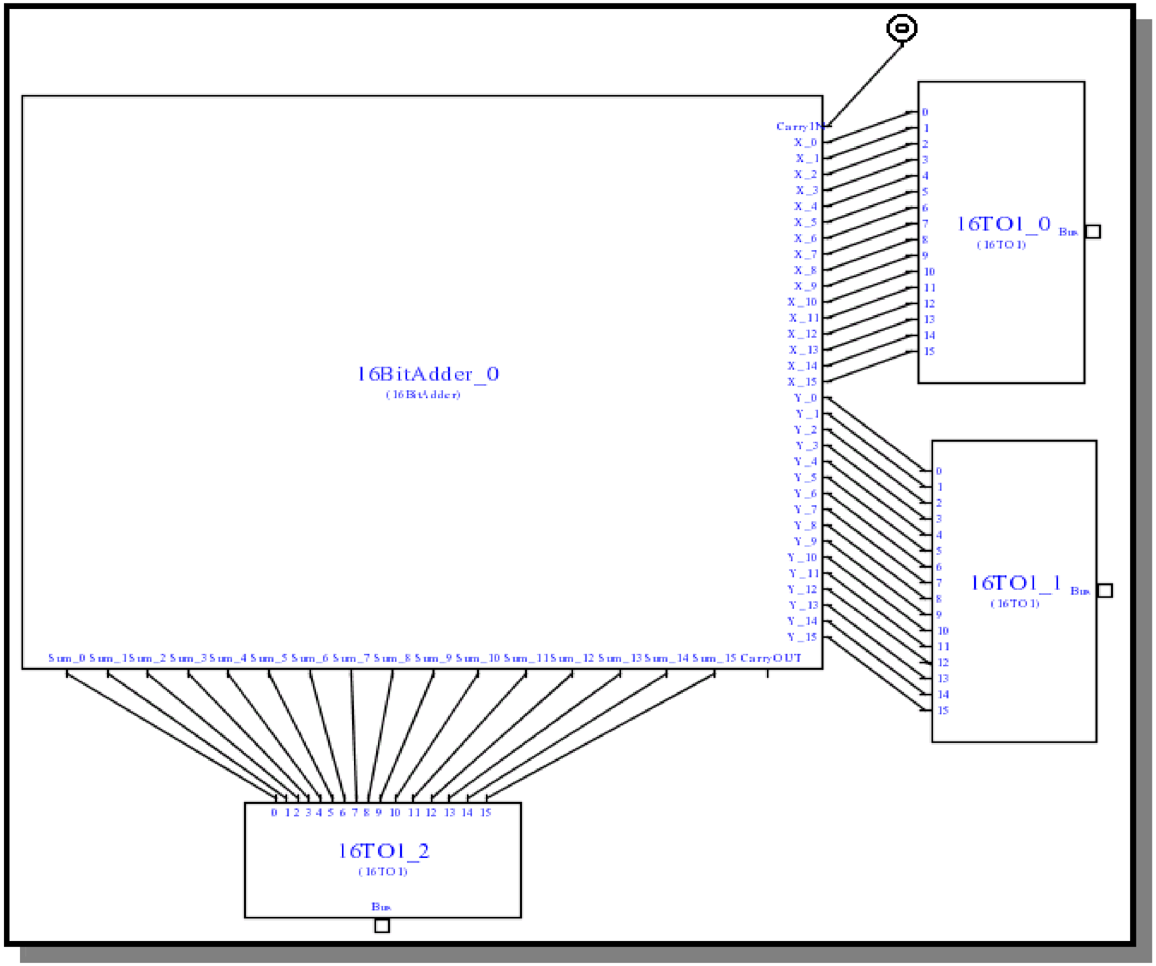


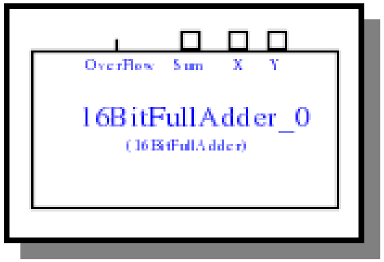



# Component Design



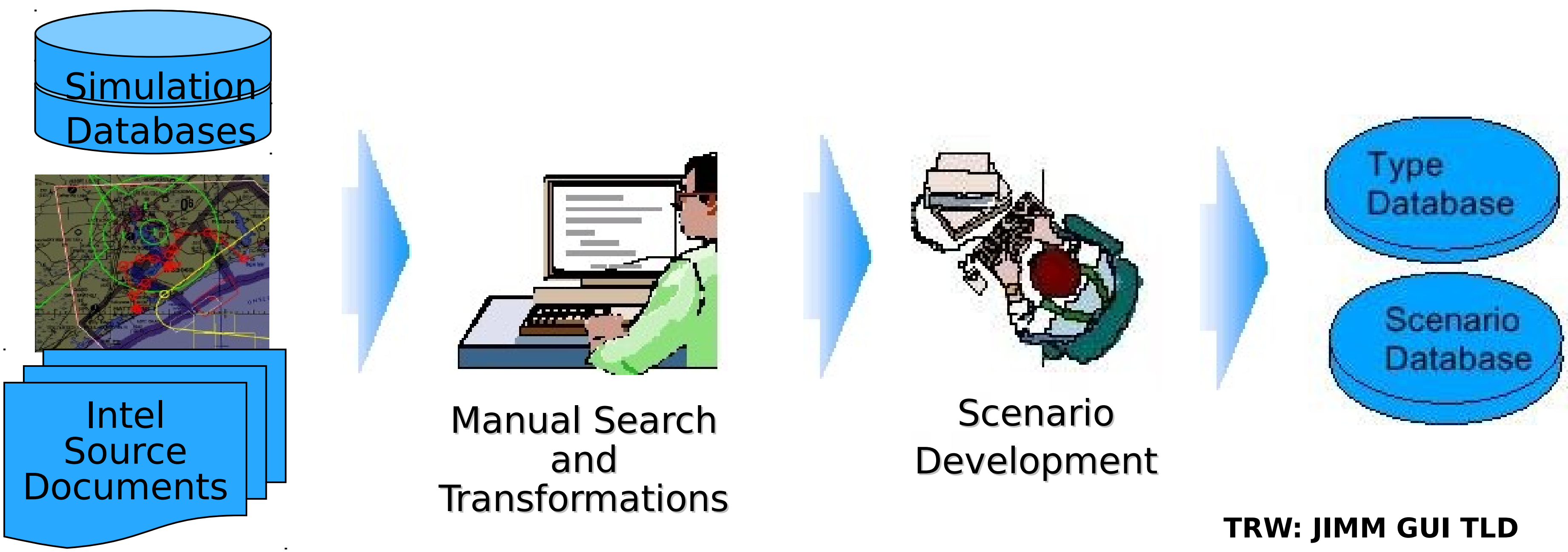
# What's Next?

- Behavior Specification!
  - Is a common representation possible?
  - Can we still check consistency of the model?
- Thinking Out Loud …
  - How do we handle temporal events?
  - State-Transition Diagrams?
  - Petri-Nets?
  - How do we make them usable by "Joe/Jane Analyst?"
  - Do these concepts apply to components created by programmers as well as those created by scenario simulation grammars?

# Intelligent Repositories

# Current Process

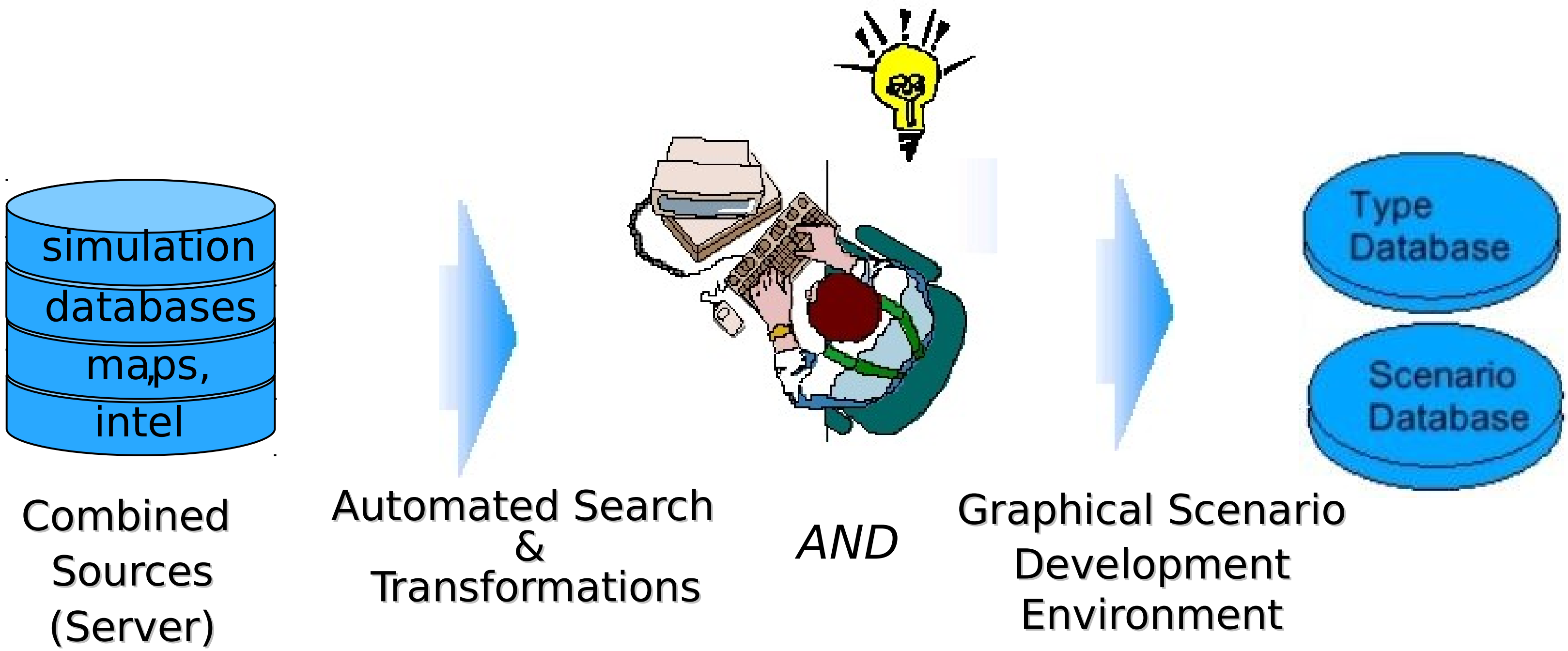
- Manual Integration of Other Models and Sources
- Reuse Limited to Same Formats
- Use “vi” in Development and Searching
- Bottom Line: Manually Intensive




TRW: JIMM GUI TLD

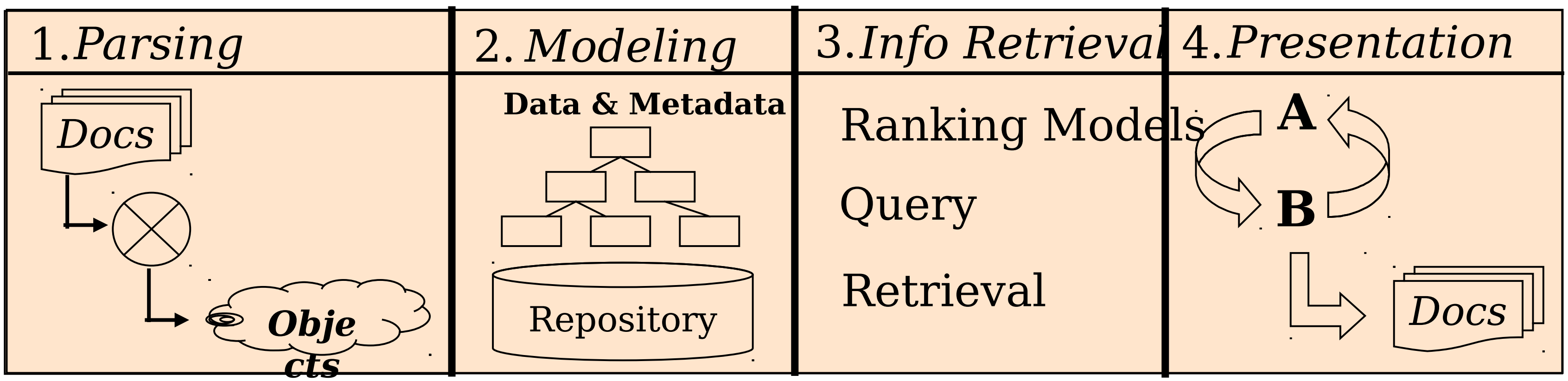
# Desired Process

- Central Repository for Simulation Entities
- Use Information Retrieval (IR) to Search
- Automatic Format Conversion
- Direct Input to Scenario Development

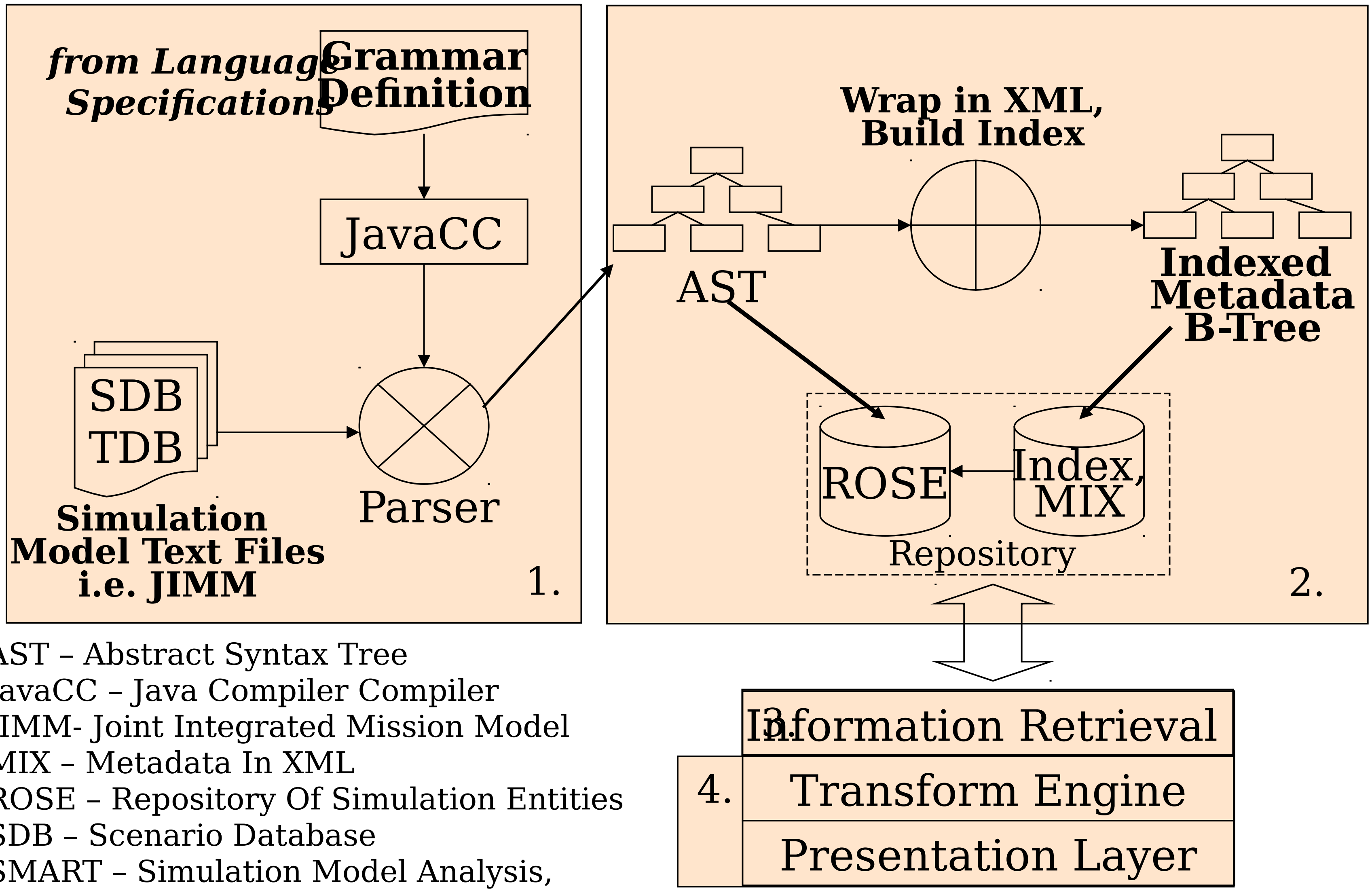


# Methodology

1. Parsing
   - Files to Java Objects
2. Modeling
   - Data and Metadata
   - Index and Repository Architecture
3. Information Retrieval
   - Scoring and Retrieval of Documents Based on Query
4. Presentation
   - Transformations, Editors, and Export Functions
   - User Interface



# Design




AST – Abstract Syntax Tree
JavaCC – Java Compiler Compiler
JIMM- Joint Integrated Mission Model
MIX – Metadata In XML
ROSE – Repository Of Simulation Entities
SDB – Scenario Database
SMART – Simulation Model Analysis, Retrieval, and Transformation
TDB – Type Database
XML – eXtensible Markup Language

# 1. Parsing

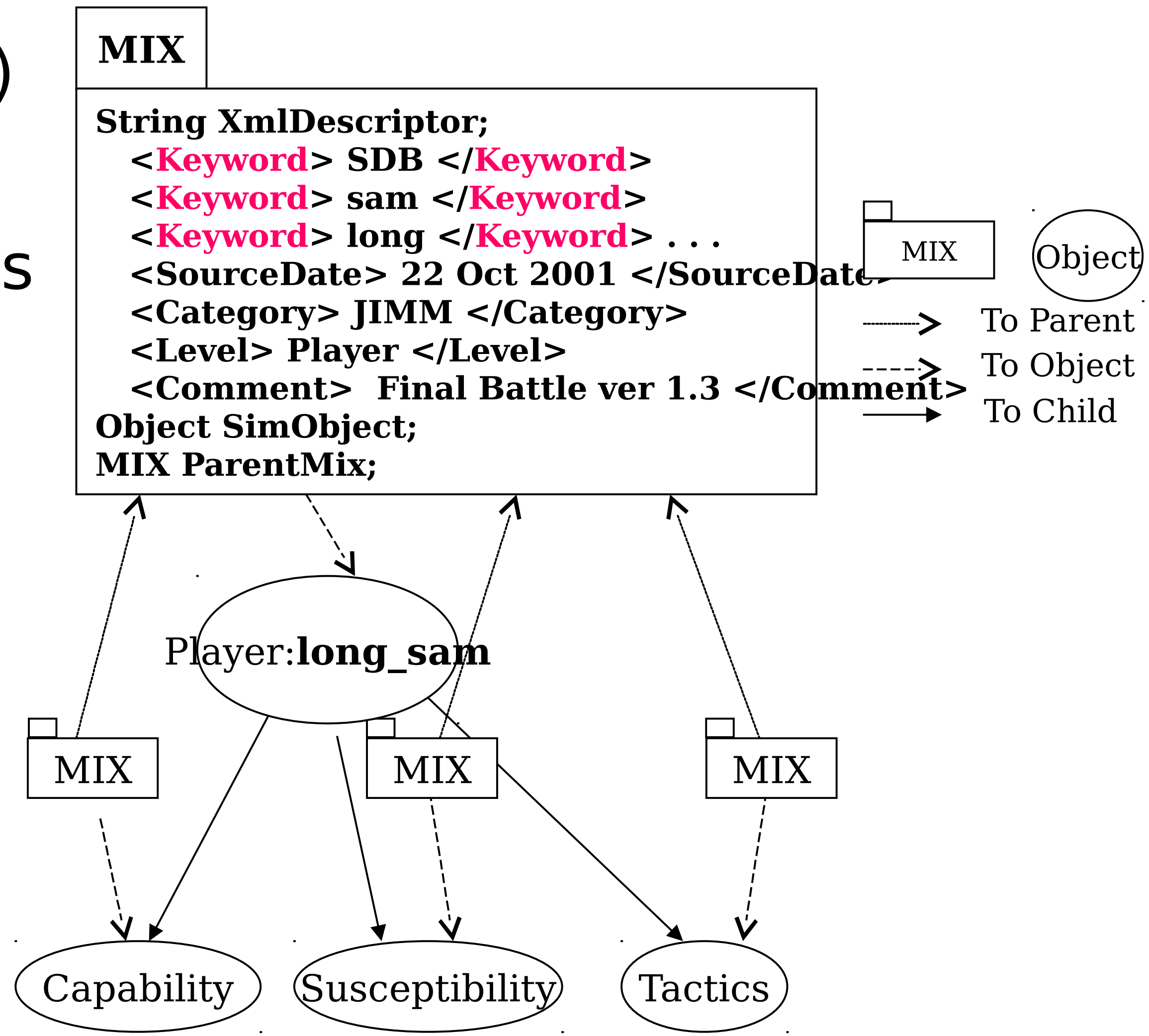
- Why?
  - Convert large text files into more manageable objects
- What’s different
  - JavaCC – Higher level of abstraction than customized parsers
- Advantages:
  - Language Specifications translate naturally to grammar definitions
  - Improves Program Understanding
  - Ease of maintenance
- Disadvantage:
  - Creates one enormous parser file →JIMM TDB *broke* JVM

# 2. Data Modeling

- Metadata:  descriptions about actual data
- Why?
  - Work with metadata instead of actual data object
  - IR models access metadata
  - Feeds index processor
- Advantages:
  - Faster
  - More Flexible
  - Source left uncontaminated
- Disadvantage:
  - Another layer of abstraction

# MIX Design

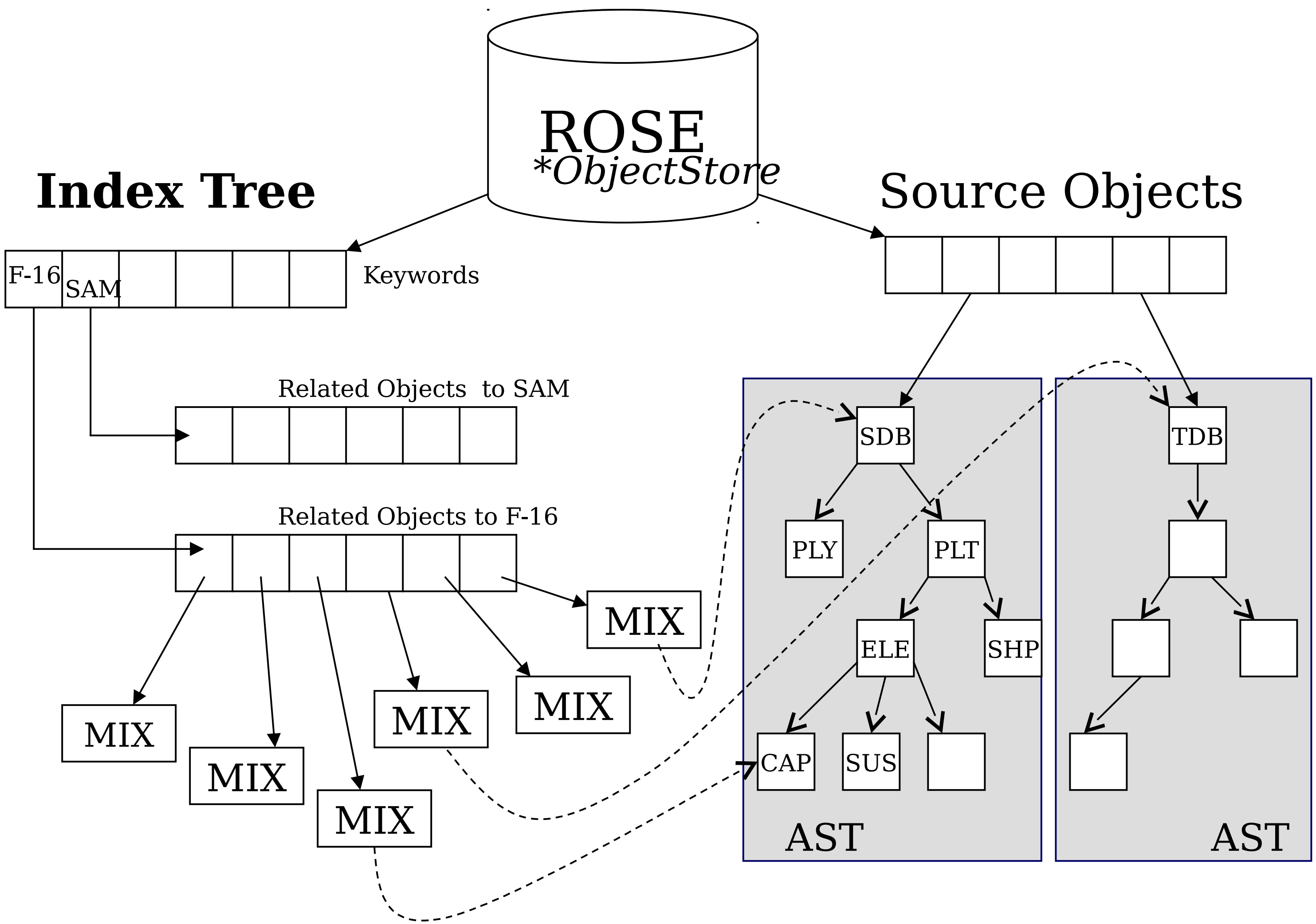
- Metadata in XML (MIX)
- Data Independent
  - User Defined Actions
- Highly Extensible
  - New Sim Model
  - New Content Handler
- Universal Data Treatment
  - IR point of view
  - Repository



# Repository

- Repository:  Storage for heterogeneous data objects
- Why?
  - Need to handle various simulation models and source documents
- Design Features
  - OODBMS
  - Single, Integrated
- Advantages:
  - OO:  Direct storage of objects; relationships maintained directly
  - DBMS:  Manages Index, Caching, Transactions, etc.
- Disadvantage:
  - Adds complexity to accommodate OODBMS

# ROSE Design




*ROSE: Repository Of Simulation Entities*

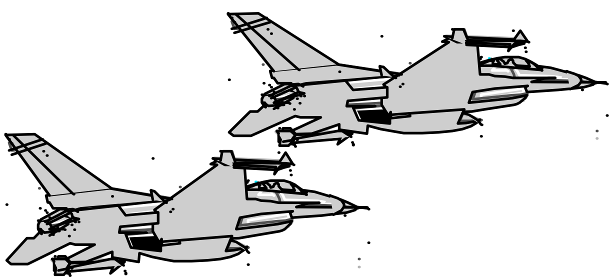
** note: not all possible lines drav*

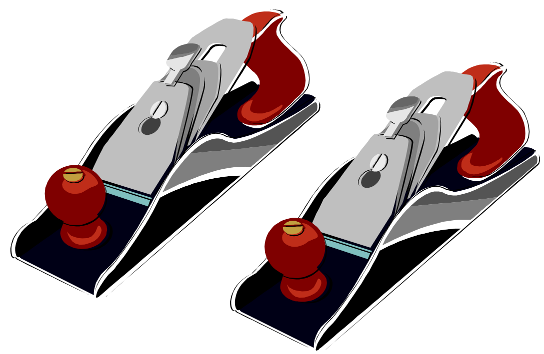
# 3. Information Retrieval

- Information Retrieval (IR):
  - Ranked list of documents relevant to query terms
  - *Similarity* score between query and document

- Data Retrieval:
  - Structured Query Language (SQL) returns unordered list of boolean rated documents (go/no-go)

Query: *jet planes*





# IR Models

- Boolean

$$sim(d_j, q) = \begin{cases} 1 & \text{if } \exists\, q_{cc} \mid (q_{cc} \in q_{dnf}) \wedge (\forall_{k_i}, g_i(d_j) = g_i(q_{cc})) \\ 0 & \text{otherwise} \end{cases}$$

- Extended Boolean

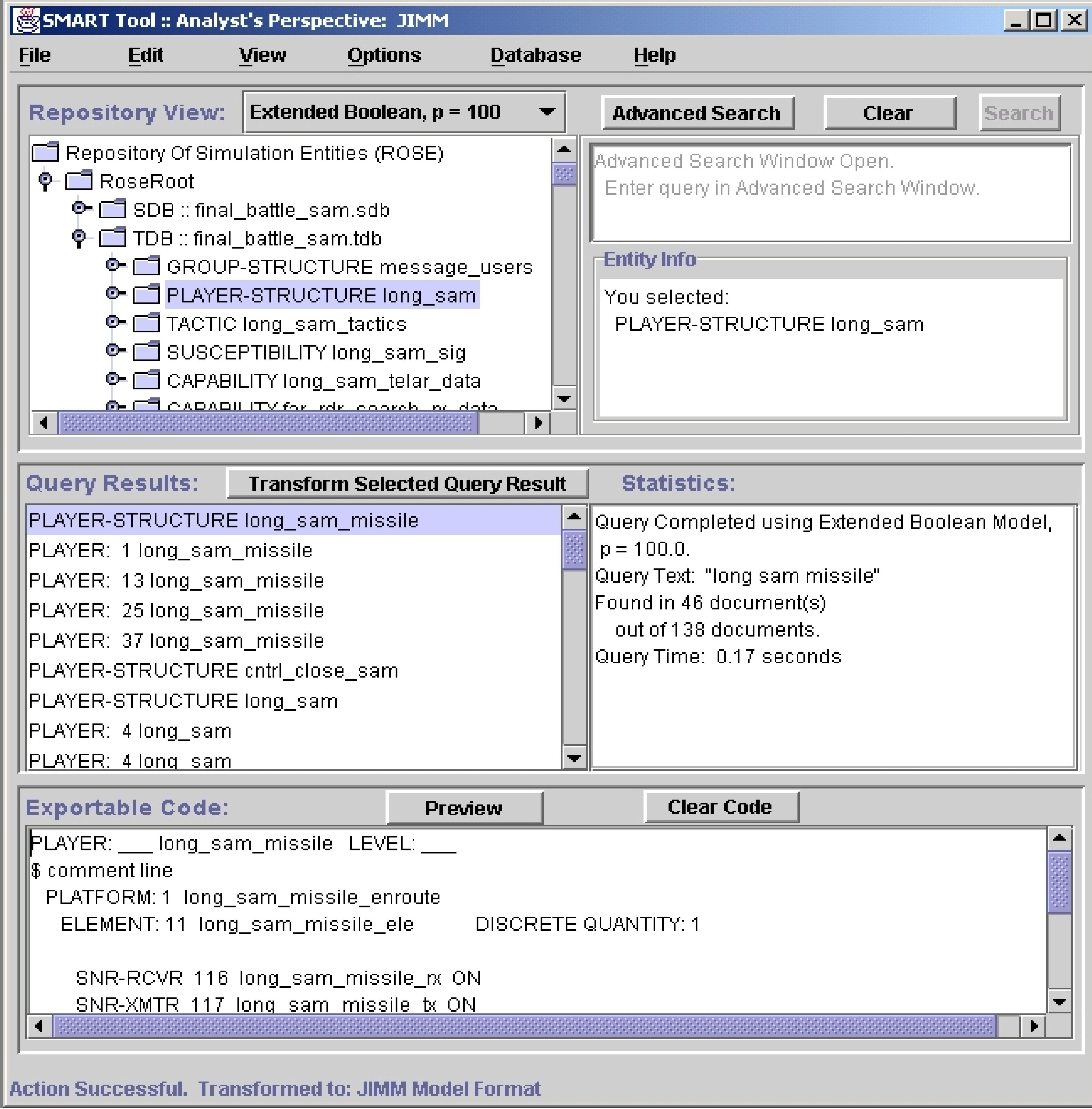
$$sim(q_{and}, d_j) = 1 - \left( \frac{(1-x_1)^p + (1-x_2)^p + \ldots + (1-x_m)^p}{m} \right)^{\frac{1}{p}}, \text{ where } x_i = \left( \frac{freq_{ij}}{\max_l \cdot freq_{lj}} \right) \left( \frac{idf_i}{\max_i \cdot idf_i} \right)$$

- Vector Model

$$sim(d_j, q) = \frac{\sum_{i=1}^{t} w_{ij} \times w_{iq}}{\sqrt{\sum_{i=1}^{t} w_{ij}^2} \times \sqrt{\sum_{i=1}^{t} w_{iq}^2}}, \quad \text{where } w_{i,j} = \frac{freq_{i,j}}{\max_l freq_{l,j}} \times \log \frac{N}{n_i}$$

# 4. Presentation

- Transformations
  - *$N^2$ Possibilities (* Model to *Model'* )
    - JIMM SDB to JIMM SDB
    - JIMM SDB to JIMM TDB
    - JIMM SDB to SUPPRESSOR SDB
    - and so on…
- Visitor Design Pattern
  - Incremental Development
  - Customizable


SMART Tool :: Analyst's Perspective: JIMM
File
Edit
View
Options
Database
Help
Repository View:
Extended Boolean, p = 100
Advanced Search
Clear
Search
Repository Of Simulation Entities (ROSE)
RoseRoot
SDB :: final_battle_sam.sdb
TDB :: final_battle_sam.tdb
GROUP-STRUCTURE message_users
PLAYER-STRUCTURE long_sam
TACTIC long_sam_tactics
SUSCEPTIBILITY long_sam_sig
CAPABILITY long_sam_telar_data
Advanced Search Window Open.
Enter query in Advanced Search Window.
Entity Info
You selected:
PLAYER-STRUCTURE long_sam
Query Results:
Transform Selected Query Result
Statistics:
PLAYER-STRUCTURE long_sam_missile
PLAYER: 1 long_sam_missile
PLAYER: 13 long_sam_missile
PLAYER: 25 long_sam_missile
PLAYER: 37 long_sam_missile
PLAYER-STRUCTURE cntrl_close_sam
PLAYER-STRUCTURE long_sam
PLAYER: 4 long_sam
PLAYER: 4 long_sam
Query Completed using Extended Boolean Model,
p = 100.0.
Query Text: "long sam missile"
Found in 46 document(s)
out of 138 documents.
Query Time: 0.17 seconds
Exportable Code:
Preview
Clear Code
PLAYER: ___ long_sam_missile LEVEL: ___
$ comment line
PLATFORM: 1 long_sam_missile_enroute
ELEMENT: 11 long_sam_missile_ele DISCRETE QUANTITY: 1
SNR-RCVR 116 long_sam_missile_rx ON
SNR-XMTR 117 long_sam_missile_tx ON
Action Successful. Transformed to: JIMM Model Format

# What's Next?

- Unification of repository technology with visual language tool
- Transitioning to non-simulation work
  - Imagery collections for ATR research
  - Imagery collections for intelligence communities
- Areas to investigate:
  - Distributed, collaborative repositories
  - Agent technologies
  - New IR models
  - Storing non-text components (binary executable components)
  - Transform techniques for other information sources

# Questions

# Background Slides

# IR Model Modifications

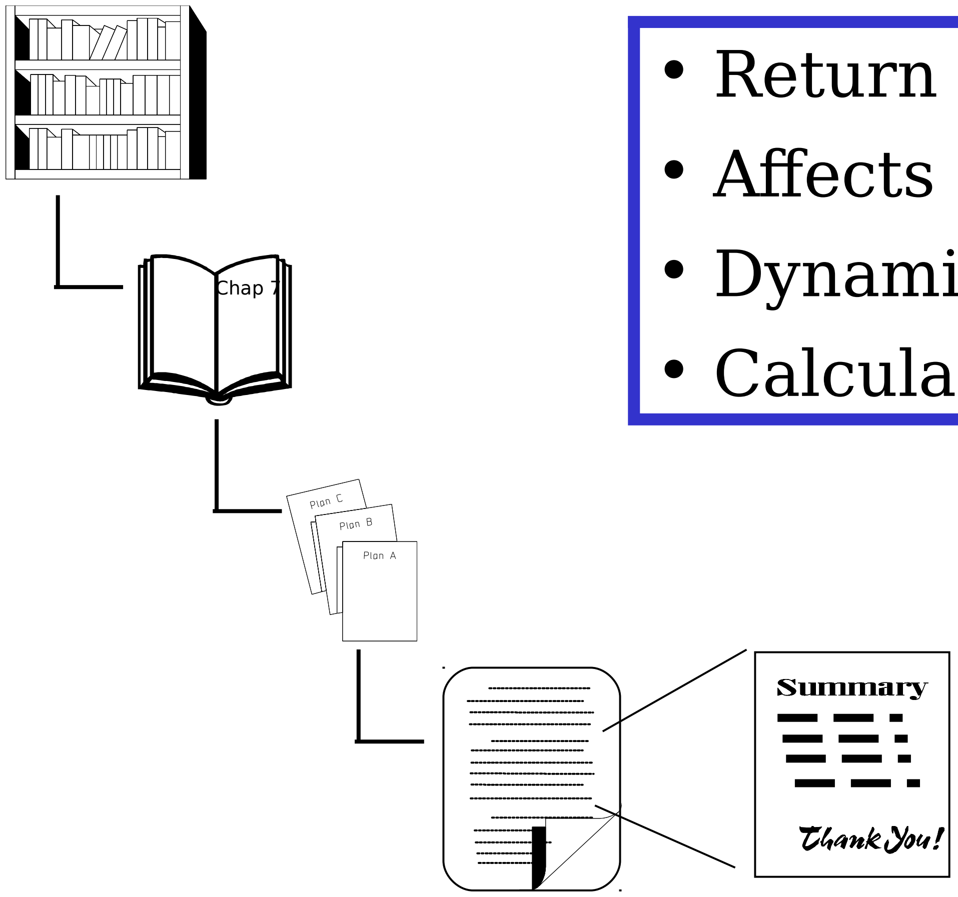
A. Dynamic Document Level (DDL) Algorithm
   - Takes Advantage of Hierarchical Structure

B. Global vs Local Analysis
   - Normalization Factors
   - Affects Score and Query Execution Time

C. Analyst’s Point of View Factor
   - Better Interpret User’s Query

# A. Dynamic Document Level

- Books →Chapters → Paragraphs →Sentences → Words

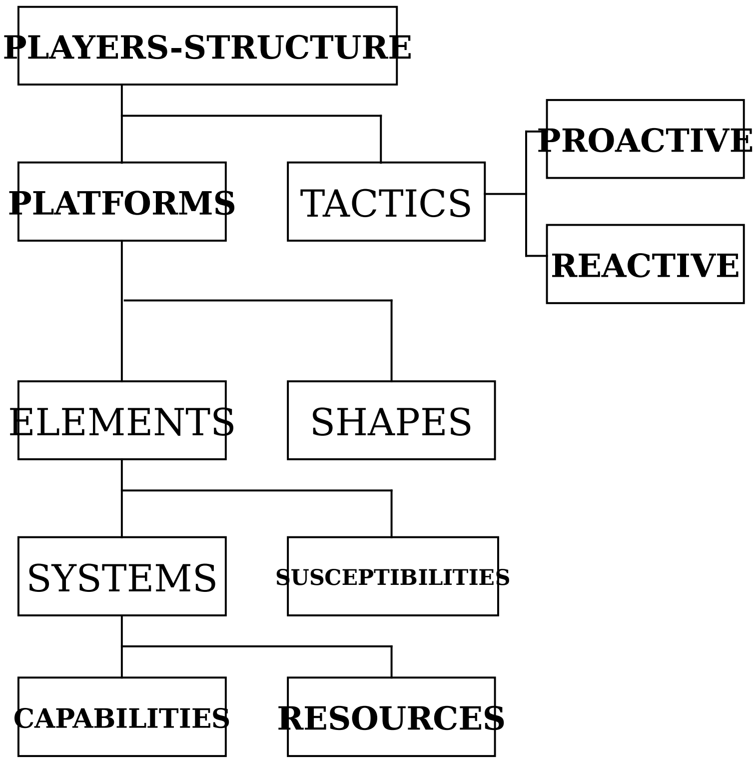


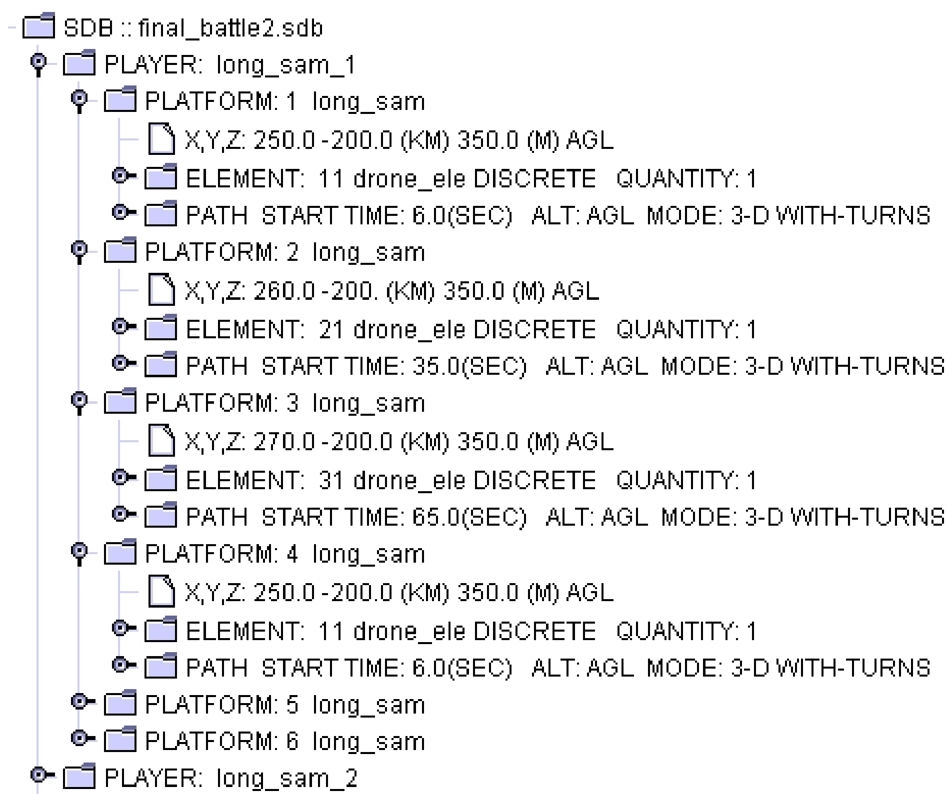

- Return results at each granularity level
- Affects how term weights are calculated
- Dynamically apply IR Models
- Calculated on the fly per query submitt

# A. Dynamic Document Level

## *JIMM HIERARCHICAL STRUCTURE*

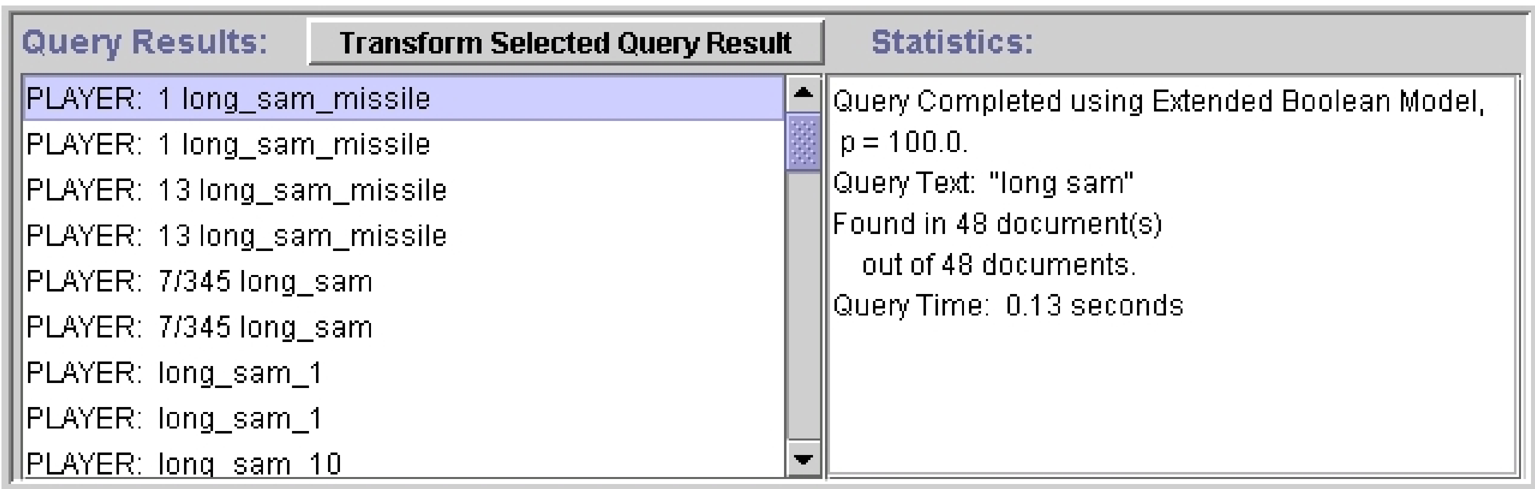
- (S)TDB → Players → Platforms → Elements …



# A. DDL Benefits

- Returned documents isolated to analyst's needs
  - e.g. Player only returned
- Increased Precision rates
  - Eliminates "noise"
  - Specific level returned



# B. Global vs Local Analysis

- Global – Documents in the entire collection
- Local – Documents retrieved from query terms
- Max $Freq_{l,j}$
  - Global
    - Terms NOT in query affects document score
  - Local
    - More correct interpretation
    - Slight additional overhead to query execution time

- Max $IDF_i$
  - Global and Local results in same relative ranking
  - Local
    - Faster Query Execution Time

# C. Point of View Factor

- Analyst's Point of View Factor
  - Biased Ranking - towards desired simulation model
  - Based on relative similarity of simulation models
- Preliminary Factors Used

| | JIMM | SUPPRESSOR | MSFD |
|---|---|---|---|
| JIMM | 1.0 | 0.8 | 0.6 |
| SUPPRESSOR | 0.8 | 1.0 | 0.6 |
| MSFD | 0.6 | 0.6 | 1.0 |

# IR Evaluations

- Based on Precision at 100% Recall
  - Precision – number of non-relevant documents retrieved
  - Recall – number of relevant documents retrieved
  - Precision at 100% Recall – how much "noise" in system

- Extended Boolean performed the best

Query*: player: long_sam radio* at No Preference

| IR Model | Precision at 100% Recall (%) | | Relevant Documents |
|---|---|---|---|
| | NP = 675 | Player = 44 | PLAYER: 4 long_sam ; score : 0.3071<br>*PLAYER-STRUCTURE long_sam; score : 0.3052<br>PLAYER: 9 long_sam ; score : 0.2893<br>PLAYER: 6/345 long_sam ; score : 0.2474<br>PLAYER: 7/345 long_sam ; score : 0.2474<br><br>* TDB player level |
| Boolean | 4 / 9 = 44 | 5 / 7 = 71 | |
| EB, p = 2 | 44 | 5 / 5 = 100 | |
| EB, p =100 | 44 | 5 / 5 = 100 | |
| Vector | 44 | 5 / 19 = 26 | |

# IR Evaluations

- Based on Precision at 100% Recall
  - Precision – number of non-relevant documents retrieved
  - Recall – number of relevant documents retrieved
  - Precision at 100% Recall – how much “noise” in system

- Extended Boolean performed the best

Query*: long_sam radio* at Player Level

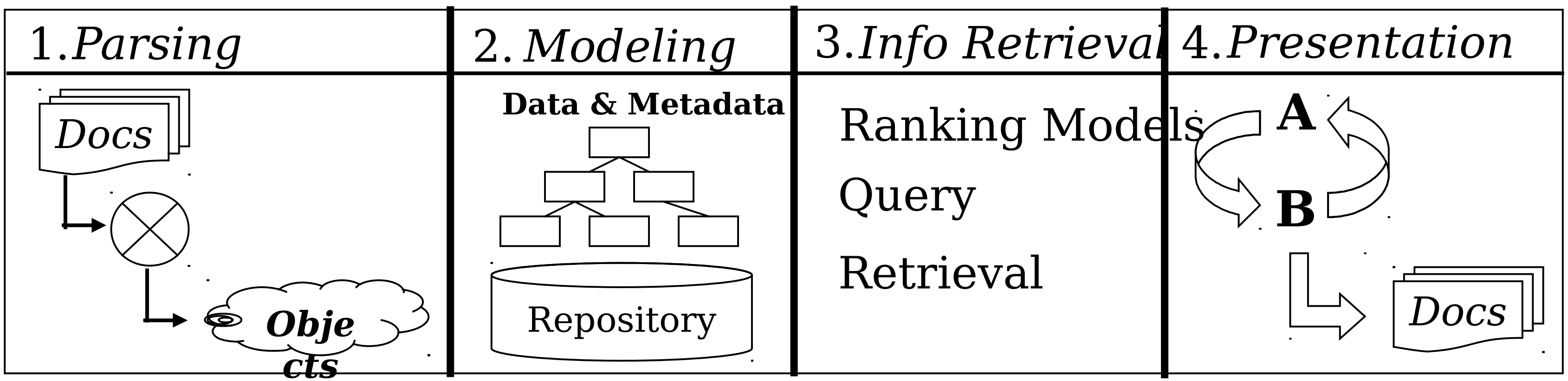
| IR Model | Precision at 100% Recall (%) | | Relevant Documents |
|---|---|---|---|
| | NP = 675 | Player = 44 | PLAYER: 4 long_sam ; score : 0.3071<br>*PLAYER-STRUCTURE long_sam; score : 0.3052<br>PLAYER: 9 long_sam ; score : 0.2893<br>PLAYER: 6/345 long_sam ; score : 0.2474<br>PLAYER: 7/345 long_sam ; score : 0.2474<br><br>* TDB player level |
| Boolean | 4 / 9 = 44 | 5 / 7 = 71 | |
| EB, p = 2 | 44 | 5 / 5 = 100 | |
| EB, p =100 | 44 | 5 / 5 = 100 | |
| Vector | 44 | 5 / 19 = 26 | |

# What’s New?

- Problem Domain
  - IR for Combat Simulation Databases
  - IR for Structured Data
- Combined Modern Tools
  - JavaCC → Parser
  - XML → Metadata
  - ObjectStore OODBMS → Repository
  - Java → Portable Program and User Interface

# What Else is New?

- Extensible Methodology
  - Parser Development and Maintenance
  - Metadata and Index Processing for Storage
  - Versatile Repository Architecture
  - Information Retrieval Process
  - Transformations



# Impact

- Big Step Towards the Ultimate Process
  - Extensible Methodology – defined steps for parsing, indexing, repository, IR, and transforms
  - Scalable Architecture – accommodate other simulation models and improvements
  - Modular Design – ease of maintenance and expansion
  - Usable Tool to Assist Current Needs
- Solid Foundation for Future Research

# Future Work

- Research Areas in Information Retrieval
  - User Relevance Feedback
  - Other IR Models, i.e. Probabilistic
  - Query Expansion Techniques

- Extend SMART System
  - Incorporate Other Simulation Models
    - Develop Additional Transformations
    - Adjust IR Ranking Models
  - Distributed and Multi-user Support
    - Convert to Server of Simulation Information